

Can survival be estimated from the yellowtail tag-recapture database?
A preliminary analysis.

Tony Wood and Steve Cadrin

NOAA/UMass CMER Program
NRC Research Associate Program

DRAFT

This information is distributed solely for the purpose of predissemination peer review. It has not been formally disseminated by NOAA. It does not represent any final agency determination of policy.

Introduction

For background information on the yellowtail flounder tagging project and database see WP3E.

A variety of techniques for estimating the survival of closed or open populations exist (e.g., Ricker 1975; Seber 2002), and mark-recapture studies have proven to be among the better methods (Brownie et al. 1985; Lebrenton et al. 1992; Quinn and Deriso 1999). With the advancement of computer technology and programming capabilities, a wide range of computer intensive methods have become widely available. One such program is MARK (White and Burnham 1999; Cooch and White 2004), which provides a means for estimating parameters from tagged animals when they are recaptured at a later date. The basic input data for the program is the reencounter history for each animal, and it uses these data to compute model parameters via numerical maximum likelihood methods (Cooch and White 2004). The mark-recapture data can be entered into the program via two methods, an encounter history file and a ‘classic’ recovery matrix, and numerous customized models can be fit to the data to estimate survival and recovery rates.

The purpose of this preliminary analysis was to examine the yellowtail mark-recapture data to see if reliable estimates of survival from the data are possible. The analyses, results and estimates are from exploratory models, and future work will be aimed at refining these models to better represent the nature of the populations and the fisheries involved.

Methods

Traditional Brownie recovery models were fit to yellowtail tag-recapture data from the yellowtail flounder tagging study carried out by the NMFS NEFSC. A preliminary examination of basic time-variant and invariant models was carried out using program MARK (White and Burnham 1999). Program MARK allows the user to fit a series time-variant tagging models to estimate either time-dependent or constant survival and recovery rates. There are four major assumptions underlying the Brownie models examined:

1. The probabilities of recapture and survival are the same for all marked animals;
2. The tagging and recapture are both instantaneous processes relative to the occasion;
3. That tagged cohorts are thoroughly mixed;
4. There is no tag loss and marks are not missed when recaptures occur (Brownie et al. 1985).

The observed yellowtail recovery (r) data were entered into the program in a triangular recovery matrix format (Brownie et al. 1985; White and Burnham 1999) with $I = 51$ months of tagging data and $J = 50$ months of recovery data:

$$r = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1J} \\ & r_{22} & \cdots & r_{2J} \\ & & \ddots & \vdots \\ & & & r_{IJ} \end{bmatrix}$$

Where r_{ij} is the recovery in year j of animals tagged in year i . To apply the models to the observed data MARK constructs matrices of expected values based on the parameters to be estimated. For example, for a tagging study with I years of tagging data and J years of recovery data the fully time dependent model is represented by the matrix (Brownie et al. 1985):

$$E_r = \begin{bmatrix} N_1 f_1 & N_2 S_2 f_2 & \cdots & N_1 (S_1 \cdots S_{J-1}) f_J \\ & N_2 f_2 & \cdots & N_2 (S_2 \cdots S_{J-1}) f_J \\ & & \ddots & \vdots \\ & & & N_I f_J \end{bmatrix},$$

where N_i is the number of individuals tagged in year i , f_i is the tag recovery rate in year i , and S_i is the survival rate in year i . Each row in the expected matrix is represented with a multinomial likelihood and the product of all row likelihoods is maximized to generate parameter estimates.

In this study, dead recovery models were used with two different parameterizations. A classic “Brownie’s parameterization” was used to fit the models (Brownie et al. 1985), and the reduced parameterization (Seber 1970; Andersen et al. 1985) was used to run model simulations. Brownie’s parameterization estimates survival (S) and recovery rate (f), where f is a product of the probability that the animal is killed (K), retrieved (α), and reported (λ) (Figure 4.2). The reduced parameter model estimates survival (S) and recovery rate (r) and can be linked algebraically to the fully parameterized model with the equation:

$$f_i = r_i(1 - S_i)$$

(Brownie et al. 1985; White and Burnham 1999; Cooch and White 2004). In this reduced parameterization r_i represents the recovery rate in year i , and differs from Brownie's parameterization in that it does not consider the source of mortality for recoveries.

The entire tag-recapture dataset was examined with four models exhibiting both time-dependent and constant survival as well as time-dependent and constant recovery rate. Three different time period specifications for the parameters were examined separately: Annual, seasonal, and monthly. Akaike's information criterion (AIC) was used to rank and select the model that achieved an optimal balance between the precision of the model and the overall fit, where precision decreases as the number of parameters in the model increases. Fit was judged with the model likelihood (L):

$$AIC = -2 \ln(L) + 2P$$

where P is equal to the number of parameters. Program MARK modifies this equation to account for differences in effective sample size (N) and lack of fit (c):

$$QAIC_c = \frac{-2 \ln(L)}{c} + 2P + \frac{2P(P+1)}{N-P-1}$$

This equation is known as the quasi likelihood adjusted AIC (Cooch and White 2004). A parametric bootstrap and median \hat{c} goodness of fit (GOF) test were carried out to determine how well the general model (fully time-dependent) fit the data. Each of these simulations used the data to generate a variance inflation factor (\hat{c}) for the reduced parameter general model. If the model fit the data perfectly a \hat{c} value of 1.0 would be generated. Deviations of \hat{c} above or below 1.0 indicate over or under-dispersion, respectively, and a \hat{c} value >3.0 often indicates a lack of fit for the model (Lebrenton et al. 1992; Cooch and White 2004). These GOF tests are available within the framework of MARK and an average value of \hat{c} from the two methods was used to adjust the $QAIC_c$ of the four models.

To quantify the differences in support between models, an index using normalized Akaike weights (w) was also calculated for each model (i) (Buckland et al. 1997):

$$w_i = \frac{e^{\frac{-\Delta QAIC_i}{2}}}{\sum e^{\frac{-\Delta QAIC_i}{2}}}$$

Results and Discussion

Both the annual and seasonal models that were fit to the data did not return acceptable diagnostics for model fit and will not be addressed further. The results from the 4 monthly models are outlined below.

The simulations to test the goodness of fit of the general model suggest a borderline model fit. The variance inflation factor (c-hat) estimates from the bootstrap goodness of fit (3.90) and the median c-hat approach (3.02) indicate overdispersion in the model (Figure 1 a and b). The mean estimate from the two methods was 3.46, which is above the designated value of 3.0 for a good model fit. In addition, the survival parameter estimates for the general model vary greatly, approaching both boundaries of 0 and 1.0 which suggests sparseness in areas of the dataset. The residual plot from the general model does not exhibit any noticeable patterns (Figure 2). However, a comprehensive examination of the residual matrix still needs to be carried out. It is likely that non-mixing is taking place for newly released individuals which could be affecting the model fit and resulting parameter estimates.

It is apparent that before estimates of either parameter can be examined further a reworking of the data and general model is necessary. For a preliminary model we have achieved a fairly good fit and it is believed that once the models and the data are refined the overall fit and the resulting parameter estimates will be greatly improved.

Table 1. Preliminary model results for 4 basic time variant and time invariant models.

Model	QAICc	Delta QAICc	QAICc Weights	Model Likelihood	Num. Par	QDeviance
{S(t) f(t)}	12046.07	0.00	1	1	75	543.30
{S(.) f(t)}	12118.21	72.14	0	0	52	661.57
{S(t) f(.)}	12287.10	241.03	0	0	51	832.46
{S(.) f(.)}	12579.23	533.16	0	0	2	1222.71

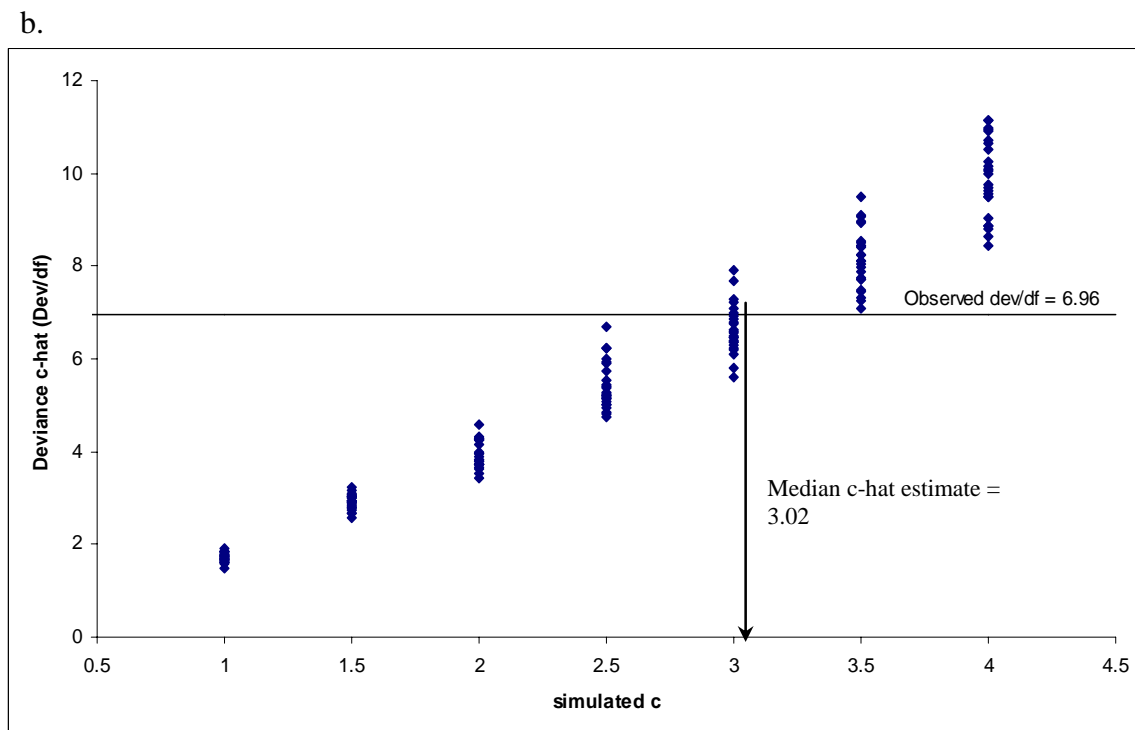
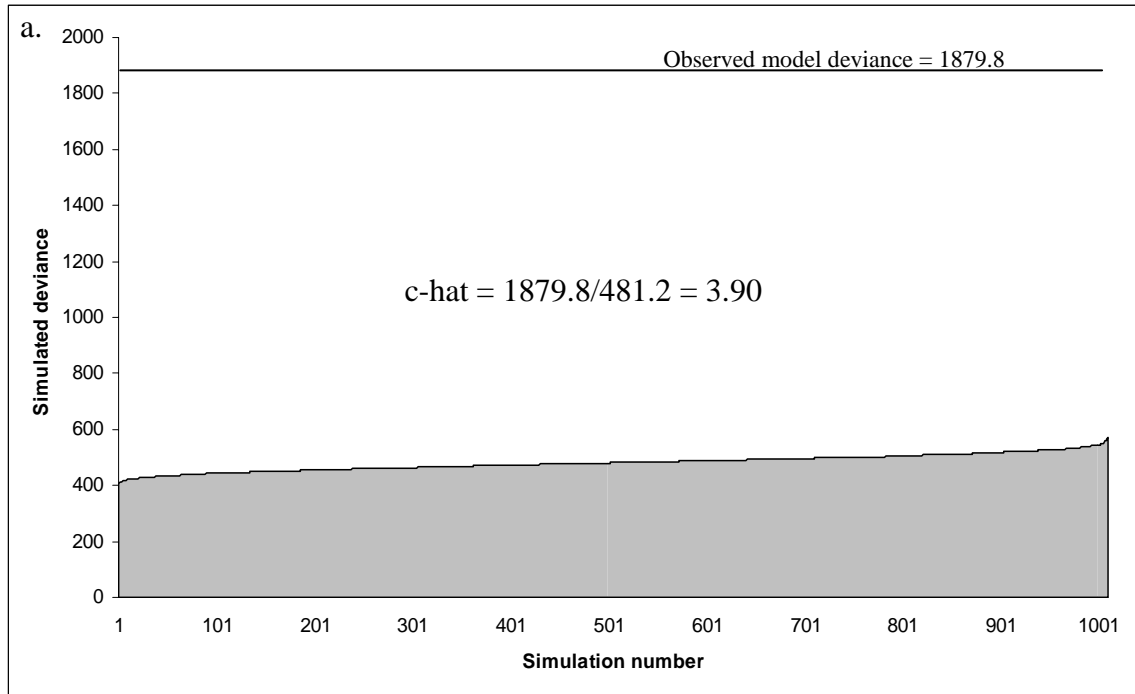


Figure 1. Results from model simulations used to estimate \hat{c} for model adjustments. a. Simulated deviances from 1000 simulations of the general (time variant) model. The estimate of \hat{c} is calculated by dividing the observed general model deviance (1879.8) by the mean of the simulated deviances (481.2). b. Simulated deviance \hat{c} (model deviance/degrees of freedom) for a range of simulated c values. The estimate of \hat{c} , determined through logistic regression, is the point where 50% of the simulated values fall above and 50% fall below the observed deviance \hat{c} for the general model (3.54).

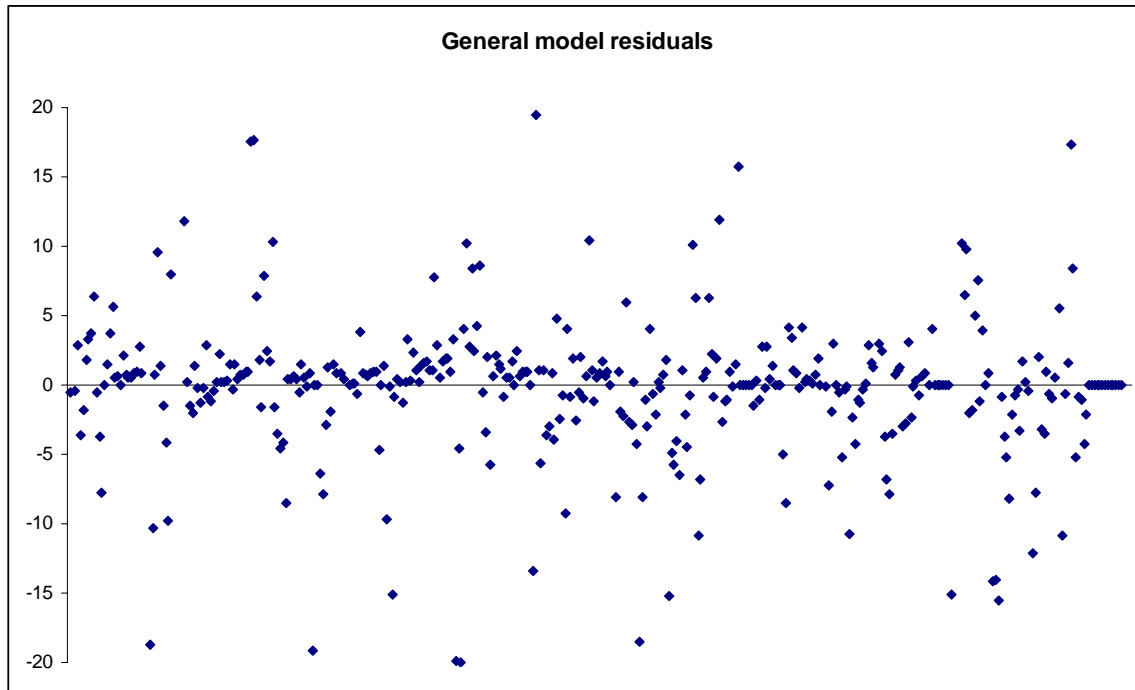


Figure 2. Residuals from the time general model with time-variant parameters. * Some large outlying residuals not shown.

Future direction

- Reparamaterize the Brownie models using information on tag reporting, tag shedding, and tag induced mortality.

- Utilize ancillary information (e.g., fishing effort) to better parameterize recovery rate and reduced the number of parameter estimates (eg. can recovery rate be reduced to a monthly cycle so that only 12 estimates of f are generated?).

References

- Brownie, C., Andersen, D. R., Burnham, K. P., & Robson, S. (1985). Statistical inference from band recovery data. A handbook 2nd Edition. *U.S Fish and Wildlife Service Resource Publications* 217.
- Buckland, S. T., Burnham, K. P., Augustin, N. H. (1997). Model selection: An integral part of inference. *Biometrics*. 53, 603–618.
- Cooch, E. & White, G. (2004). *Program MARK “A Gentle Introduction.”* 5th Edn. <http://www.phidot.org/software/mark/docs/book/>, 545 pp.
- Lebrenton, J. D., Burnham, K. P., Clobert, J. & Andersen, D. R. (1992). Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monographs* 62, 67-118.
- Ricker, W.E. (1975). Computation and interpretation of biological statistics of fish populations. *Bulletin of the Fisheries Research Board of Canada* 191, 1-381.
- Seber, G. A. F. (1970). Estimating time-specific survival and reporting rates for adult birds from band returns. *Biometrika* 57, 313-318.
- Seber, G. A. F. (2002). *The estimation of animal abundance and related parameters.* 2nd edn. New York: MacMillan, 654 pp.
- White, G. C. & Burnham, K. P.(1999). Program MARK: Survival estimation from populations of marked animals. *Bird Study* 46 Supplement, 120-138.